



RAW SEQUENCE LISTING ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/031,496B
Source: 1/FWJ6 -
Date Processed by STIC: 3/12/04 -

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION AND PATENTIN SOFTWARE QUESTIONS, PLEASE CONTACT MARK SPENCER, TELEPHONE: 703-308-4212; FAX: 703-308-4221

Effective 12/13/03: TELEPHONE: 571-272-2510; FAX: 571-273-0221

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 4.1 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker/chkr41note.htm>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for UK disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebc/cfs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
3. Hand Carry directly to (EFFECTIVE 12/01/03):
U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office, Box Sequence, Room 4B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

Revised 10/08/03

Raw Sequence Listing Error Summary

ERROR DETECTED

SUGGESTED CORRECTION

SERIAL NUMBER:

10/031,496B

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 Wrapped Nucleics
 Wrapped Aminos The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 2 Invalid Line Length The rules require that a line not exceed 72 characters in length. This includes white spaces.
- 3 Misaligned Amino
 Numbering The numbering under each 5th amino acid is misaligned. Do not use tab codes between numbers; use **space characters**, instead.
- 4 Non-ASCII The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
- 5 Variable Length Sequence(s) contain n's or Xaa's representing more than one residue. **Per Sequence Rules, each n or Xaa can only represent a single residue.** Please present the **maximum** number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6 PatentIn 2.0
 "bug" A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for **Artificial or Unknown sequences.**
- 7 Skipped Sequences
 (OLD RULES) Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence:
 (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
 (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading).
 (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
 This sequence is intentionally skipped

 Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
- 8 Skipped Sequences
 (NEW RULES) Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence.
 <210> sequence id number
 <400> sequence id number
 000
- 9 Use of n's or Xaa's
 (NEW RULES) Use of n's and/or Xaa's have been detected in the Sequence Listing.
 Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.
 In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 10 Invalid <213>
 Response Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
- 11 Use of <220> Sequence(s) missing the <220> "Feature" and associated numeric identifiers and responses.
 Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.
 (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12 PatentIn 2.0
 "bug" Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
- 13 Misuse of n/Xaa "n" can **only** represent a single nucleotide; "Xaa" can only represent a single amino acid



IFW16

RAW SEQUENCE LISTING

DATE: 03/12/2004

PATENT APPLICATION: US/10/031,496B

TIME: 14:56:14

Input Set : A:\NREL 99-45.ST25.txt

Output Set: N:\CRF4\03122004\J031496B.raw

3 <110> APPLICANT: National Renewable Energy Laboratory
 5 <120> TITLE OF INVENTION: Cellobiohydrolase I Gene and Improved Variants
 7 <130> FILE REFERENCE: NREL 99-45
 9 <140> CURRENT APPLICATION NUMBER: 10/031,496B
 10 <141> CURRENT FILING DATE: 2002-01-14
 12 <160> NUMBER OF SEQ ID NOS: 120
 14 <170> SOFTWARE: PatentIn version 3.2
 16 <210> SEQ ID NO: 1
 17 <211> LENGTH: 28 *invalid <213> response*
 18 <212> TYPE: DNA *see item 10*
 19 <213> ORGANISM: Synthetic DNA *on Enr*
 21 <400> SEQUENCE: 1 *summary sheet*
 22 agagagtctta gacacggagc ttacaggc
 25 <210> SEQ ID NO: 2
 26 <211> LENGTH: 35
 27 <212> TYPE: DNA
 28 <213> ORGANISM: Synthetic DNA
 30 <400> SEQUENCE: 2
 31 aaagaagcgc ggcgcgcgcct gcactctcca atcgg
 34 <210> SEQ ID NO: 3
 35 <211> LENGTH: 24
 36 <212> TYPE: DNA
 37 <213> ORGANISM: Synthetic DNA
 39 <400> SEQUENCE: 3
 40 ggcggaaacc cgcctggcac cacc
 43 <210> SEQ ID NO: 4
 44 <211> LENGTH: 1550
 45 <212> TYPE: DNA
 46 <213> ORGANISM: Trichoderma reesei
 49 <220> FEATURE:
 50 <221> NAME/KEY: misc_signal
 51 <222> LOCATION: (1)..(51)
 53 <220> FEATURE:
 54 <221> NAME/KEY: CDS
 55 <222> LOCATION: (3)..(1550)
 57 <220> FEATURE:
 58 <221> NAME/KEY: misc_feature
 59 <222> LOCATION: (52)..(1344)
 61 <220> FEATURE:
 62 <221> NAME/KEY: misc_feature
 63 <222> LOCATION: (1345)..(1435)
 65 <220> FEATURE:
 66 <221> NAME/KEY: misc_binding

28
35
24

Please correct this error in subsequent sequences, if present.

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/031,496B

DATE: 03/12/2004

TIME: 14:56:14

Input Set : A:\NREL 99-45.ST25.txt

Output Set: N:\CRF4\03122004\J031496B.raw

```

67 <222> LOCATION: (1436)..(1550)
69 <400> SEQUENCE: 4
70 at gta tgc gaa gtt ggc cgt cat ctc ggc ctt ctt ggc cac agc tgc      47
71   Val Ser Glu Val Gly Arg His Leu Gly Leu Leu Gly His Ser Ser
72   1           5           10           15
74 tgc tca gtc ggc ctg cac tct cca atc gga gac tca ccc gcc tct gac      95
75 Cys Ser Val Gly Leu His Ser Pro Ile Gly Asp Ser Pro Ala Ser Asp
76           20           25           30
78 atg gca gaa atg ctc gtc tgg tgg cac gtg cac tca aca gac agg ctc      143
79 Met Ala Glu Met Leu Val Trp Trp His Val His Ser Thr Asp Arg Leu
80           35           40           45
82 cgt ggt cat cga cgc caa ctg gcg ctg gac tca cgc tac gaa cag cag      191
83 Arg Gly His Arg Arg Gln Leu Ala Leu Asp Ser Arg Tyr Glu Gln Gln
84           50           55           60
86 cac gaa ctg cta cga tgg caa cac ttg gag ctc gac cct atg tcc tga      239
87 His Glu Leu Leu Arg Trp Gln His Leu Glu Leu Asp Pro Met Ser
88           65           70           75
90 caa cga gac ctg cgc gaa gaa ctg ctg tct gga cgg tgc cgc cta cgc      287
91 Gln Arg Asp Leu Arg Glu Glu Leu Leu Ser Gly Arg Cys Arg Leu Arg
92           80           85           90
94 gtc cac gta cgg agt tac cac gag cgg taa cag cct ctc cat tgg ctt      335
95 Val His Val Arg Ser Tyr His Glu Arg           Gln Pro Leu His Trp Leu
96 95           100           105
98 tgt cac cca gtc tgc gca gaa gaa cgt tgg cgc tgc cct tta cct tat      383
99 Cys His Pro Val Cys Ala Glu Glu Arg Trp Arg Ser Pro Leu Pro Tyr
100 110           115           120           125
102 ggc gag cga cac gac cta cca gga att cac cct gct tgg caa cga gtt      431
103 Gly Glu Arg His Asp Leu Pro Gly Ile His Pro Ala Trp Gln Arg Val
104           130           135           140
106 ctc ttt cga tgt tga tgt ttc gca gct gcc gtg cgg ctt gaa cgg agc      479
107 Leu Phe Arg Cys           Cys Phe Ala Ala Ala Val Arg Leu Glu Arg Ser
108           145           150           155
110 tct cta ctt cgt gtc cat gga cgc gga tgg cgt gag caa gta tcc      527
111 Ser Leu Leu Arg Val His Gly Arg Gly Trp Trp Arg Glu Gln Val Ser
112           160           165           170
114 cac caa cac cgc tgg cgc caa gta cgg cac ggg gta ctg tga cag cca      575
115 His Gln His Arg Trp Arg Gln Val Arg His Gly Val Leu           Gln Pro
116           175           180           185
118 gtg tcc ccg cga tct gaa gtt cat caa tgg cca ggc caa cgt tga ggg      623
119 Val Ser Pro Arg Ser Glu Val His Gln Trp Pro Gly Gln Arg           Gly
120           190           195           200
122 ctg gga gcc gtc atc caa caa cgc gaa cac ggg cat tgg agg aca cgg      671
123 Leu Gly Ala Val Ile Gln Gln Arg Glu His Gly His Trp Arg Thr Arg
124           205           210           215
126 aag ctg ctg ctc tga gat gga tat ctg gga ggc caa ctc cat ctc cga      719
127 Lys Leu Leu Leu           Asp Gly Tyr Leu Gly Gly Gln Leu His Leu Arg
128           220           225           230
130 ggc tct tac ccc cca ccc ttg cac gac tgt cgg cca gga gat ctg cga      767
131 Gly Ser Tyr Pro Pro Pro Leu His Asp Cys Arg Pro Gly Asp Leu Arg

```

RAW SEQUENCE LISTING

DATE: 03/12/2004

PATENT APPLICATION: US/10/031,496B

TIME: 14:56:14

Input Set : A:\NREL 99-45.ST25.txt

Output Set: N:\CRF4\03122004\J031496B.raw

132	235	240	245	
134	ggg tga tgg gtg cgg cgg aac tta ctc cga taa cag ata tgg cgg cac	815		
135	Gly Trp Val Arg Arg Asn Leu Leu Arg Gln Ile Trp Arg His			
136	250 255 260			
138	ttg cga tcc cga tgg ctg cga ctg gaa ccc ata cgg cct ggg caa cac	863		
139	Leu Arg Ser Arg Trp Leu Arg Leu Glu Pro Ile Pro Pro Gly Gln His			
140	265 270 275			
142	cag ctt cta cgg ccc tgg ctc aag ctt tac cct cga tac cac caa gaa	911		
143	Gln Leu Leu Arg Pro Trp Leu Lys Leu Tyr Pro Arg Tyr His Gln Glu			
144	280 285 290 295			
146	att gac cgt tgt cac cca gtt cga gac gtc ggg tgc cat caa cgg ata	959		
147	Ile Asp Arg Cys His Pro Val Arg Asp Val Gly Cys His Gln Pro Ile			
148	300 305 310			
150	cta tgt cca gaa tgg cgt cac ttt cca gca gcc caa cgc cga gct tgg	1007		
151	Leu Cys Pro Glu Trp Arg His Phe Pro Ala Ala Gln Arg Arg Ala Trp			
152	315 320 325			
154	tag tta ctc tgg caa cga gct caa cga tga tta ctg cac agc tga gga	1055		
155	Leu Leu Trp Gln Arg Ala Gln Arg Leu Leu His Ser Gly			
156	330 335 340			
158	ggc aga att cgg cgg atc ctc ttt ctc aga caa ggg cgg cct gac tca	1103		
159	Gly Arg Ile Arg Arg Ile Leu Phe Leu Arg Gln Gly Arg Pro Asp Ser			
160	345 350 355			
162	gtt caa gaa ggc tac ctc tgg cgg cat ggt tct ggt cat gag tct gtg	1151		
163	Val Gln Glu Gly Tyr Leu Trp Arg His Gly Ser Gly His Glu Ser Val			
164	360 365 370			
166	gga tga tta cta cgc caa cat gct gtg gct gga ctc cac cta ccc gac	1199		
167	Gly Leu Leu Arg Gln His Ala Val Ala Gly Leu His Leu Pro Asp			
168	375 380 385			
170	aaa cga gac ctc ctc cac acc cgg tgc cgt gcg cgg aag ctg ctc cac	1247		
171	Lys Arg Asp Leu Leu His Thr Arg Cys Arg Ala Arg Lys Leu Leu His			
172	390 395 400			
174	cag ctc cgg tgt ccc tgc tca ggt cga atc tca gtc tcc caa cgc caa	1295		
175	Gln Leu Arg Cys Pro Cys Ser Gly Arg Ile Ser Val Ser Gln Arg Gln			
176	405 410 415			
178	ggt cac ctt ctc caa cat caa gtt cgg acc cat tgg cag cac cgg caa	1343		
179	Gly His Leu Leu Gln His Gln Val Arg Thr His Trp Gln His Arg Gln			
180	420 425 430 435			
182	ccc tag cgg cgg caa ccc tcc cgg cgg aaa ccc gcc tgg cac cac cac	1391		
183	Pro Arg Arg Gln Pro Ser Arg Arg Lys Pro Ala Trp His His His			
184	440 445 450			
186	cac cgg cgg ccc agc cac tac cac tgg aag ctc tcc cgg acc tac cca	1439		
187	His Pro Pro Pro Ser His Tyr His Trp Lys Leu Ser Arg Thr Tyr Pro			
188	455 460 465			
190	gtc tca cta cgg cca gtg cgg cgg tat tgg cta cag cgg ccc cac ggt	1487		
191	Val Ser Leu Arg Pro Val Arg Arg Tyr Trp Leu Gln Arg Pro His Gly			
192	470 475 480			
194	ctg cgc cag cgg cac aac ttg cca ggt cct gmc cct tac tac tct cag	1535		
195	Leu Arg Gln Arg His Asn Leu Pro Gly Pro Xaa Pro Tyr Tyr Ser Gln			
196	485 490 495			

OK

RAW SEQUENCE LISTING

DATE: 03/12/2004

PATENT APPLICATION: US/10/031,496B

TIME: 14:56:14

Input Set : A:\NREL 99-45.ST25.txt

Output Set: N:\CRF4\03122004\J031496B.raw

```

198 tgc ctg taa agc tcc                                     1550
199 Cys Leu      Ser Ser
200      500
203 <210> SEQ ID NO: 5
204 <211> LENGTH: 78
205 <212> TYPE: PRT
206 <213> ORGANISM: Trichoderma reesei
208 <400> SEQUENCE: 5
210 Val Ser Glu Val Gly Arg His Leu Gly Leu Leu Gly His Ser Ser Cys
211 1      5      10      15
214 Ser Val Gly Leu His Ser Pro Ile Gly Asp Ser Pro Ala Ser Asp Met
215      20      25      30
218 Ala Glu Met Leu Val Trp Trp His Val His Ser Thr Asp Arg Leu Arg
219      35      40      45
222 Gly His Arg Arg Gln Leu Ala Leu Asp Ser Arg Tyr Glu Gln Gln His
223      50      55      60
226 Glu Leu Leu Arg Trp Gln His Leu Glu Leu Asp Pro Met Ser
227 65      70      75
230 <210> SEQ ID NO: 6
231 <211> LENGTH: 25
232 <212> TYPE: PRT
233 <213> ORGANISM: Trichoderma reesei
235 <400> SEQUENCE: 6
237 Gln Arg Asp Leu Arg Glu Glu Leu Leu Ser Gly Arg Cys Arg Leu Arg
238 1      5      10      15
241 Val His Val Arg Ser Tyr His Glu Arg
242      20      25
245 <210> SEQ ID NO: 7
246 <211> LENGTH: 42
247 <212> TYPE: PRT
248 <213> ORGANISM: Trichoderma reesei
250 <400> SEQUENCE: 7
252 Gln Pro Leu His Trp Leu Cys His Pro Val Cys Ala Glu Glu Arg Trp
253 1      5      10      15
256 Arg Ser Pro Leu Pro Tyr Gly Glu Arg His Asp Leu Pro Gly Ile His
257      20      25      30
260 Pro Ala Trp Gln Arg Val Leu Phe Arg Cys
261      35      40
264 <210> SEQ ID NO: 8
265 <211> LENGTH: 40
266 <212> TYPE: PRT
267 <213> ORGANISM: Trichoderma reesei
269 <400> SEQUENCE: 8
271 Cys Phe Ala Ala Val Arg Leu Glu Arg Ser Ser Leu Leu Arg Val
272 1      5      10      15
275 His Gly Arg Gly Trp Trp Arg Glu Gln Val Ser His Gln His Arg Trp
276      20      25      30
279 Arg Gln Val Arg His Gly Val Leu
280      35      40

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/031,496B

DATE: 03/12/2004

TIME: 14:56:14

Input Set : A:\NREL 99-45.ST25.txt

Output Set: N:\CRF4\03122004\J031496B.raw

```

283 <210> SEQ ID NO: 9
284 <211> LENGTH: 16
285 <212> TYPE: PRT
286 <213> ORGANISM: Trichoderma reesei
288 <400> SEQUENCE: 9
290 Gln Pro Val Ser Pro Arg Ser Glu Val His Gln Trp Pro Gly Gln Arg
291 1 5 10 15
294 <210> SEQ ID NO: 10
295 <211> LENGTH: 21
296 <212> TYPE: PRT
297 <213> ORGANISM: Trichoderma reesei
299 <400> SEQUENCE: 10
301 Gly Leu Gly Ala Val Ile Gln Gln Arg Glu His Gly His Trp Arg Thr
302 1 5 10 15
305 Arg Lys Leu Leu Leu
306 20
309 <210> SEQ ID NO: 11
310 <211> LENGTH: 28
311 <212> TYPE: PRT
312 <213> ORGANISM: Trichoderma reesei
314 <400> SEQUENCE: 11
316 Asp Gly Tyr Leu Gly Gly Gln Leu His Leu Arg Gly Ser Tyr Pro Pro
317 1 5 10 15
320 Pro Leu His Asp Cys Arg Pro Gly Asp Leu Arg Gly
321 20 25
324 <210> SEQ ID NO: 12
325 <211> LENGTH: 8
326 <212> TYPE: PRT
327 <213> ORGANISM: Trichoderma reesei
329 <400> SEQUENCE: 12
331 Trp Val Arg Arg Asn Leu Leu Arg
332 1 5
335 <210> SEQ ID NO: 13
336 <211> LENGTH: 69
337 <212> TYPE: PRT
338 <213> ORGANISM: Trichoderma reesei
340 <400> SEQUENCE: 13
342 Gln Ile Trp Arg His Leu Arg Ser Arg Trp Leu Arg Leu Glu Pro Ile
343 1 5 10 15
346 Pro Pro Gly Gln His Gln Leu Leu Arg Pro Trp Leu Lys Leu Tyr Pro
347 20 25 30
350 Arg Tyr His Gln Glu Ile Asp Arg Cys His Pro Val Arg Asp Val Gly
351 35 40 45
354 Cys His Gln Pro Ile Leu Cys Pro Glu Trp Arg His Phe Pro Ala Ala
355 50 55 60
358 Gln Arg Arg Ala Trp
359 65
362 <210> SEQ ID NO: 14
363 <211> LENGTH: 8

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/031,496B

DATE: 03/12/2004
TIME: 14:56:15

Input Set : A:\NREL 99-45.ST25.txt
Output Set: N:\CRF4\03122004\J031496B.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:4; Xaa Pos. 493
Seq#:18; Xaa Pos. 57
Seq#:32; Xaa Pos. 57

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/031,496B

DATE: 03/12/2004

TIME: 14:56:15

Input Set : A:\NREL 99-45.ST25.txt

Output Set: N:\CRF4\03122004\J031496B.raw

L:195 M:258 W: Mandatory Feature missing, <223> Tag not found for SEQ ID#:4
L:195 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:1535
L:450 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:18 after pos.:48
L:702 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:32 after pos.:48